MP3 and AAC Audio Compression

Topics on Algorithms

Seok Won Lee

Contents

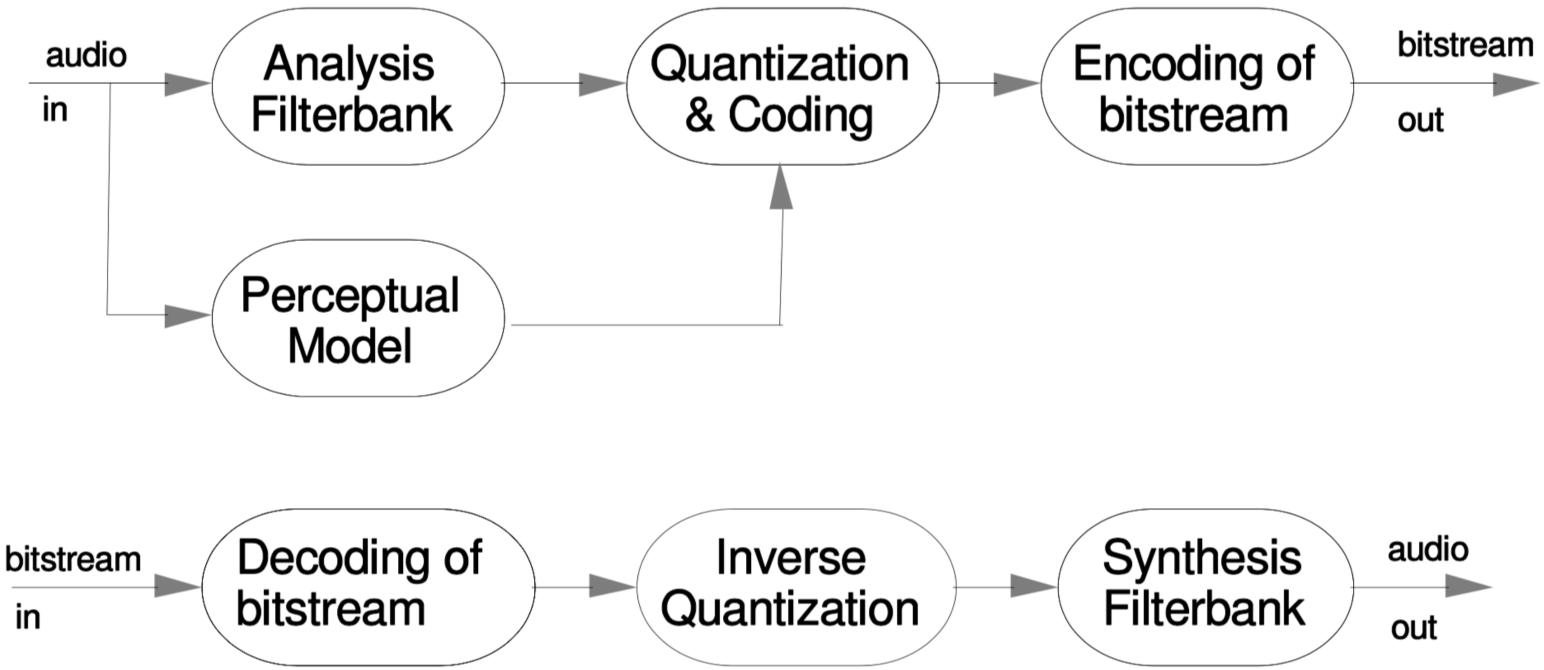
High Quality Audio Coding

- The compression is as efficient as possible, i.e. the compressed file is as small as possible.

- The reconstructed audio sounds exactly (or as close as possible) to the original audio before compression.

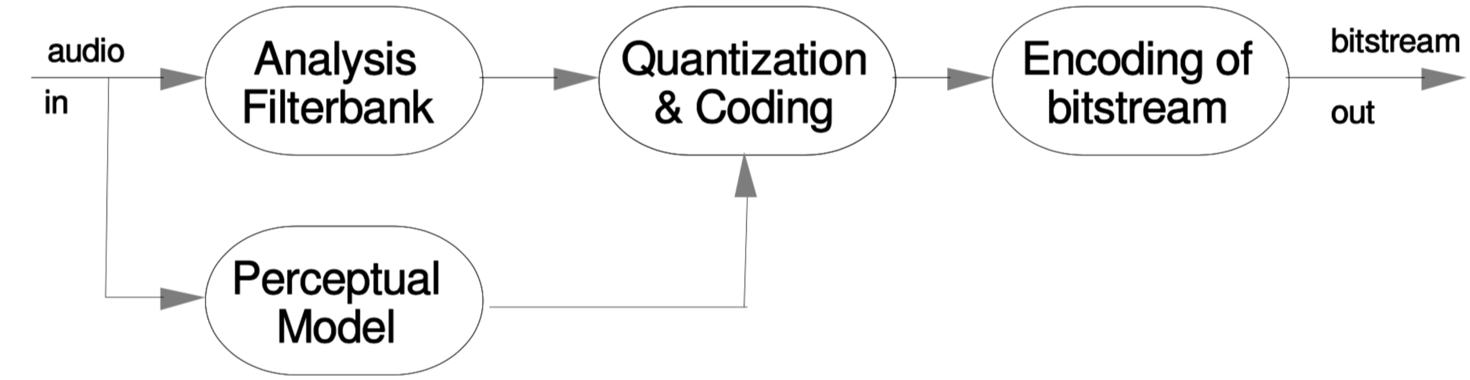
=> **Let's use knowledge from psychoacoustics!**

Perceptual Encoding



- Both **MP3(MPEG-1 Audio Layer-3)** and **AAC(MPEG-2 Advanced Audio Coding)** are examples of perceptual encoding.

Perceptual Encoding

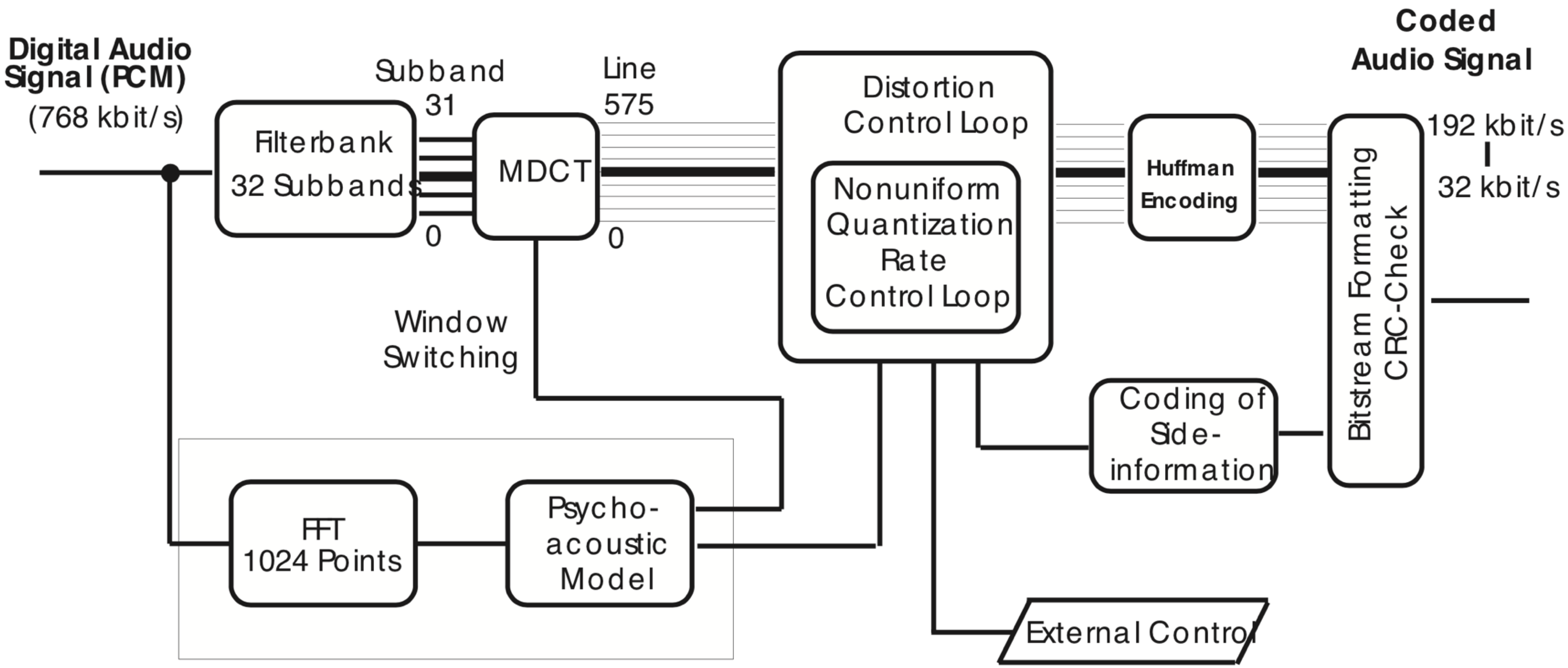


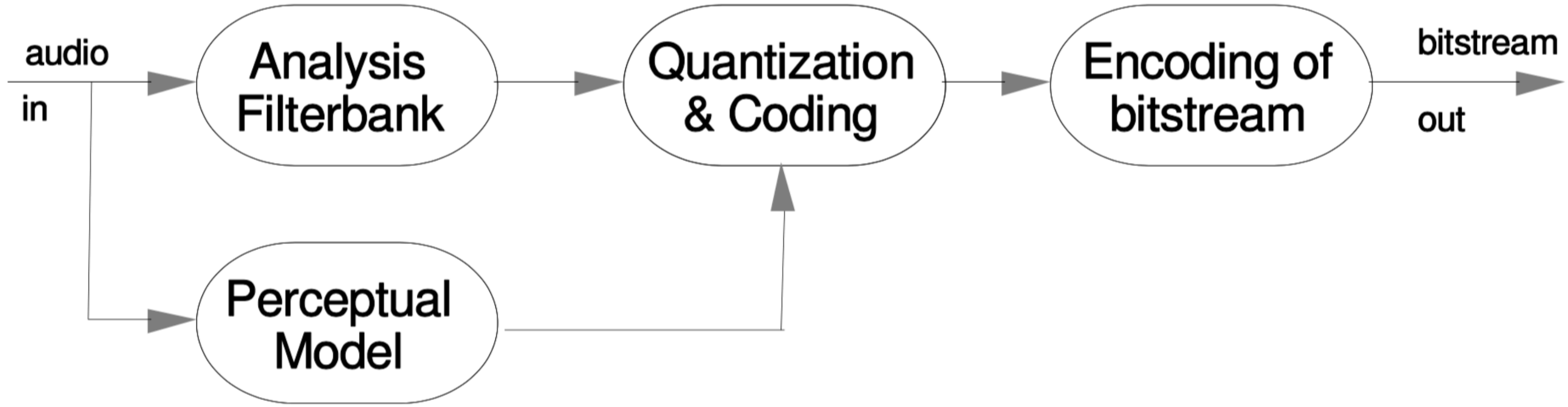
Filterbank : Decompose the input signal into subsampled spectral components (time/frequency domain)

Perceptual Model : Compute an estimate of the actual (time and frequency dependent) masking threshold using rules known from psychoacoustics

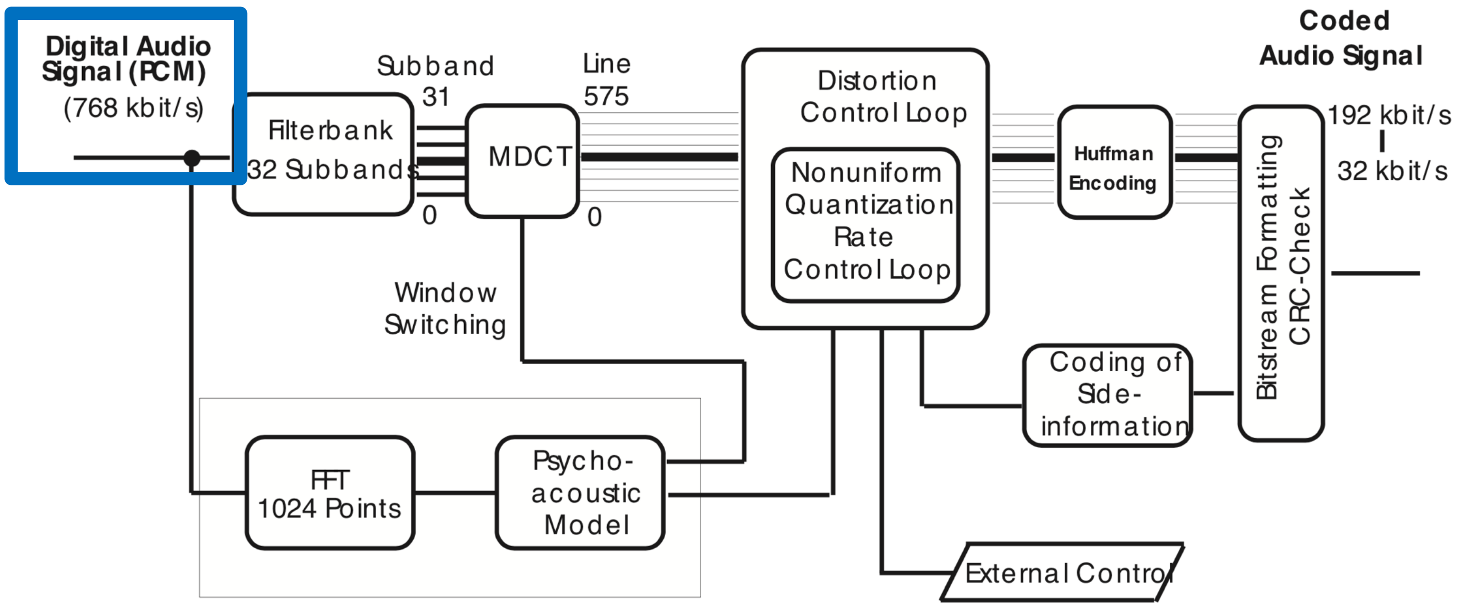
Quantization & Coding : Quantize and code the spectral components with the aim of keeping the noise below the masking threshold.

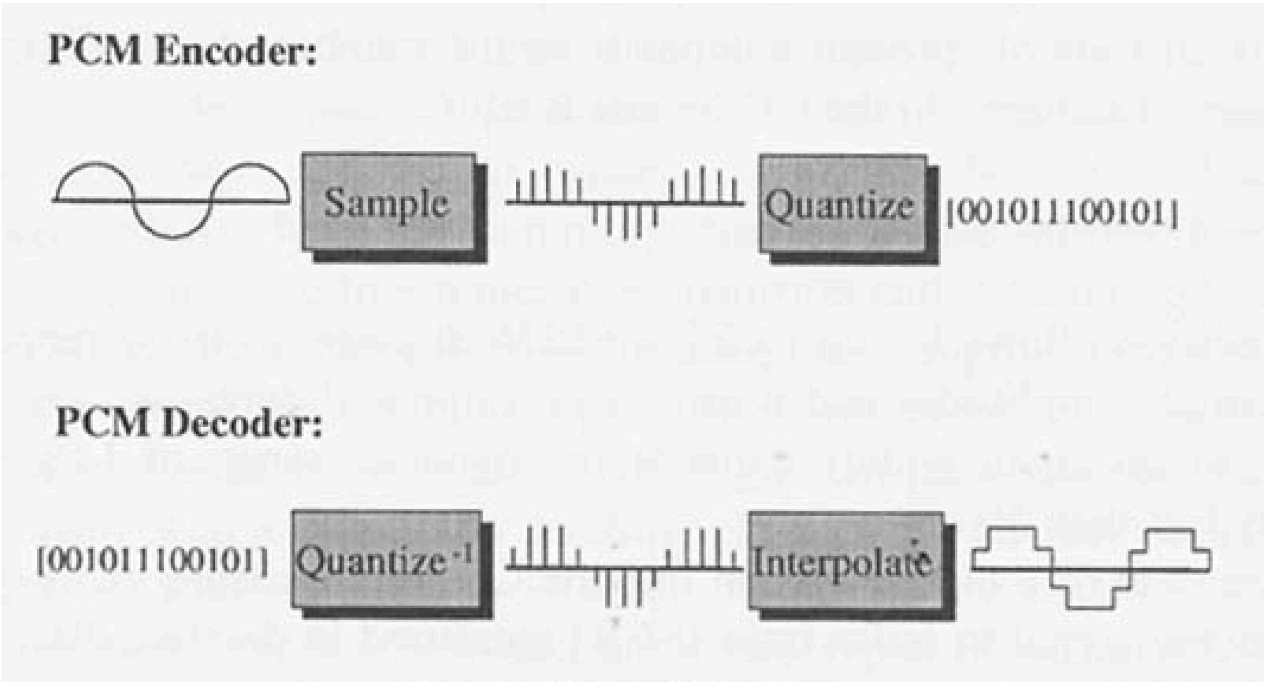
MPEG-1 Audio Layer-3 Encoder



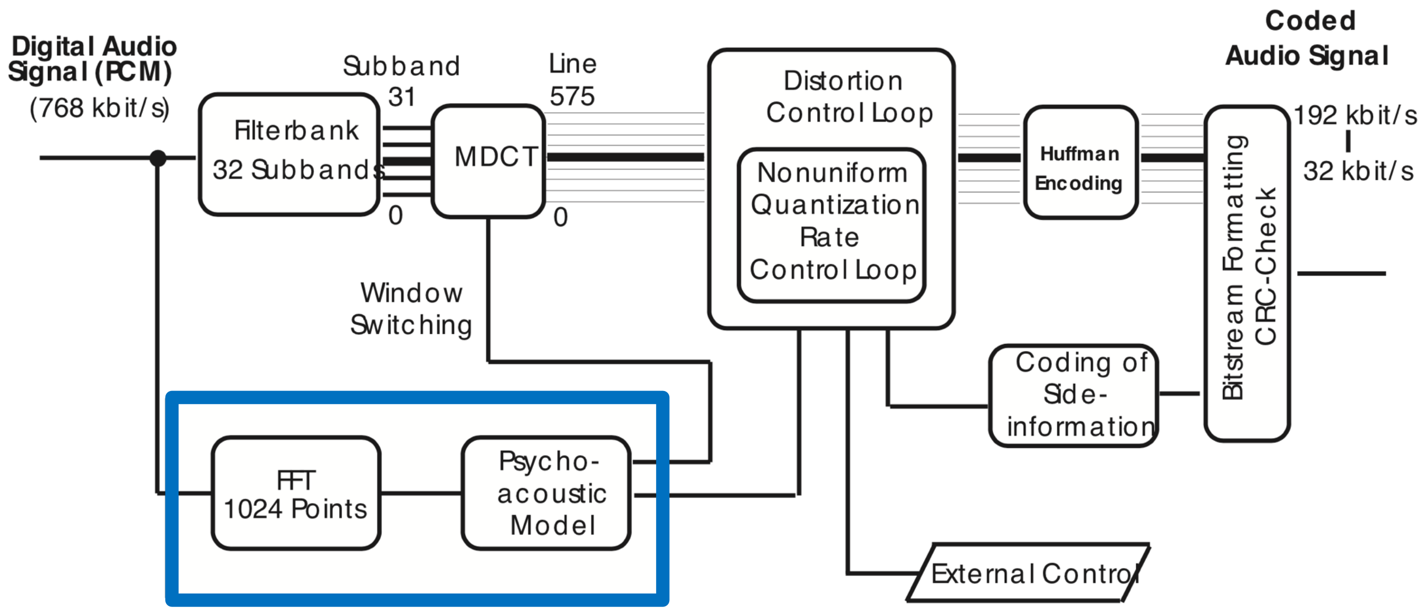


MPEG-1 Audio Layer-3 Encoder





MPEG-1 Audio Layer-3 Encoder



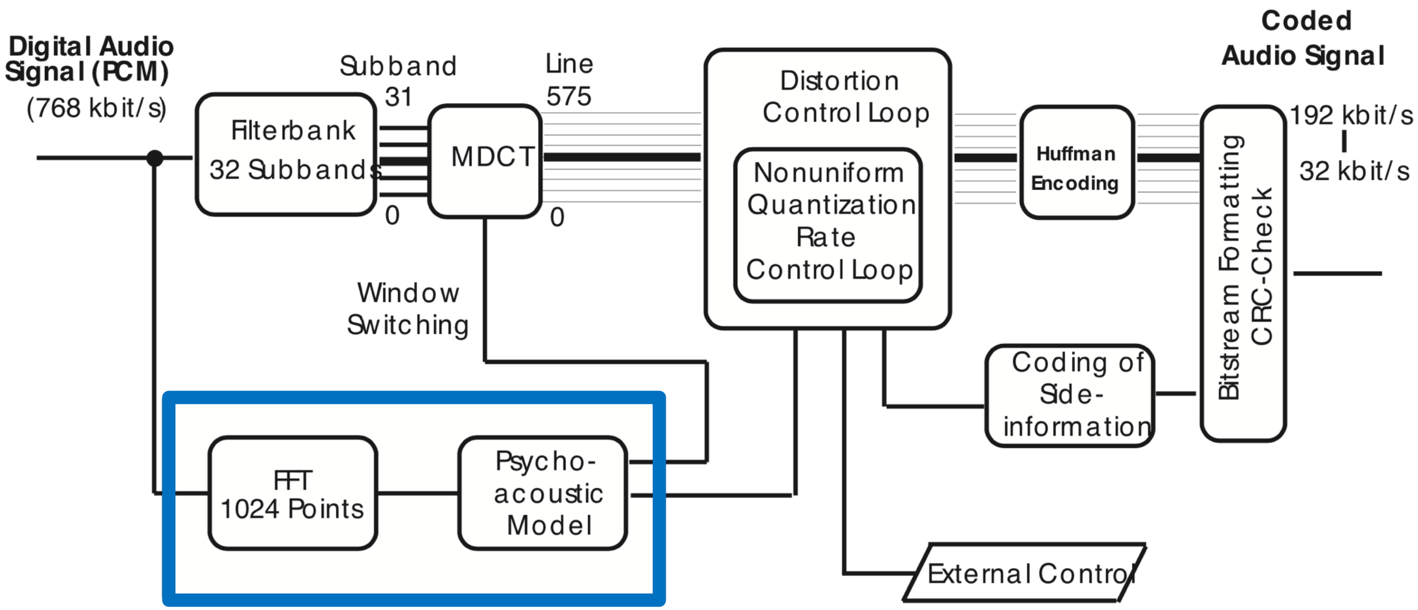
Fast Fourier Transformation (FFT)

Input : A sequence of 1152 PCM samples

Output : Evaluation of PCM samples at 1024 points

- It provides input data for psychoacoustic model.

MPEG-1 Audio Layer-3 Encoder



Psychoacoustic Model

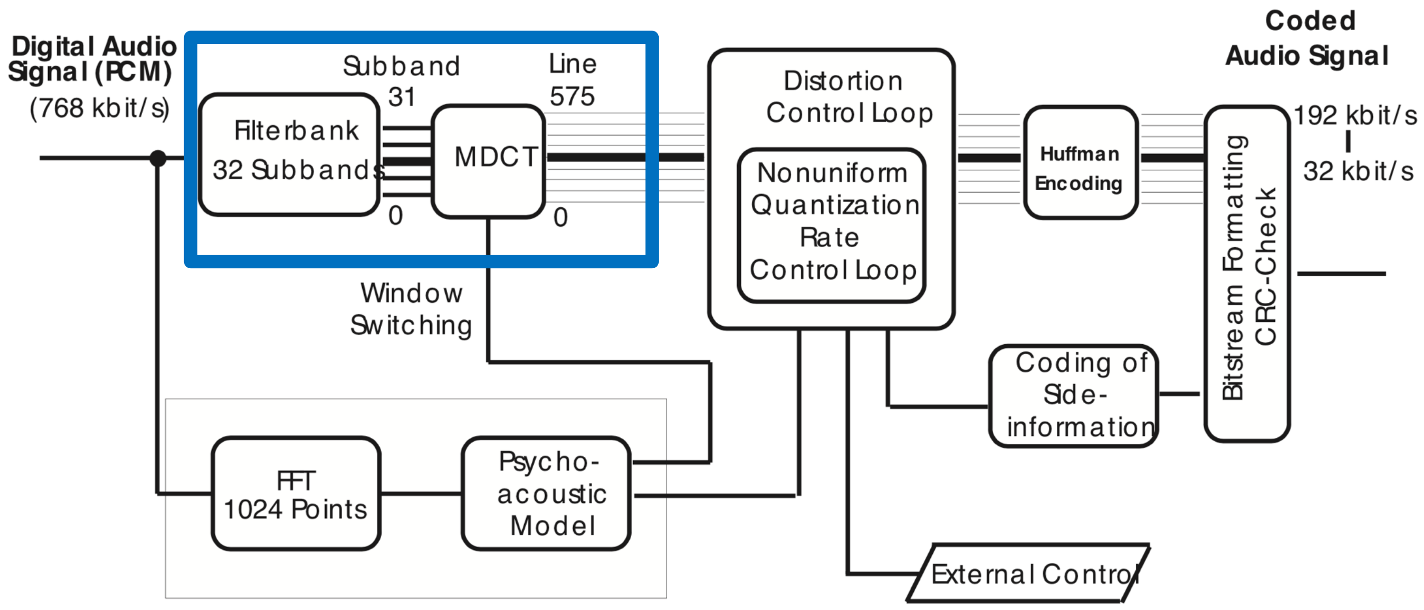
Input : Output of FFT

Output : What window function to use in MDCT

The number of scalefactor bands to use (usually 12 or 21)

The allowed noise of each scalefactor band

MPEG-1 Audio Layer-3 Encoder



Filterbank

Input : A sequence of 1152 PCM samples

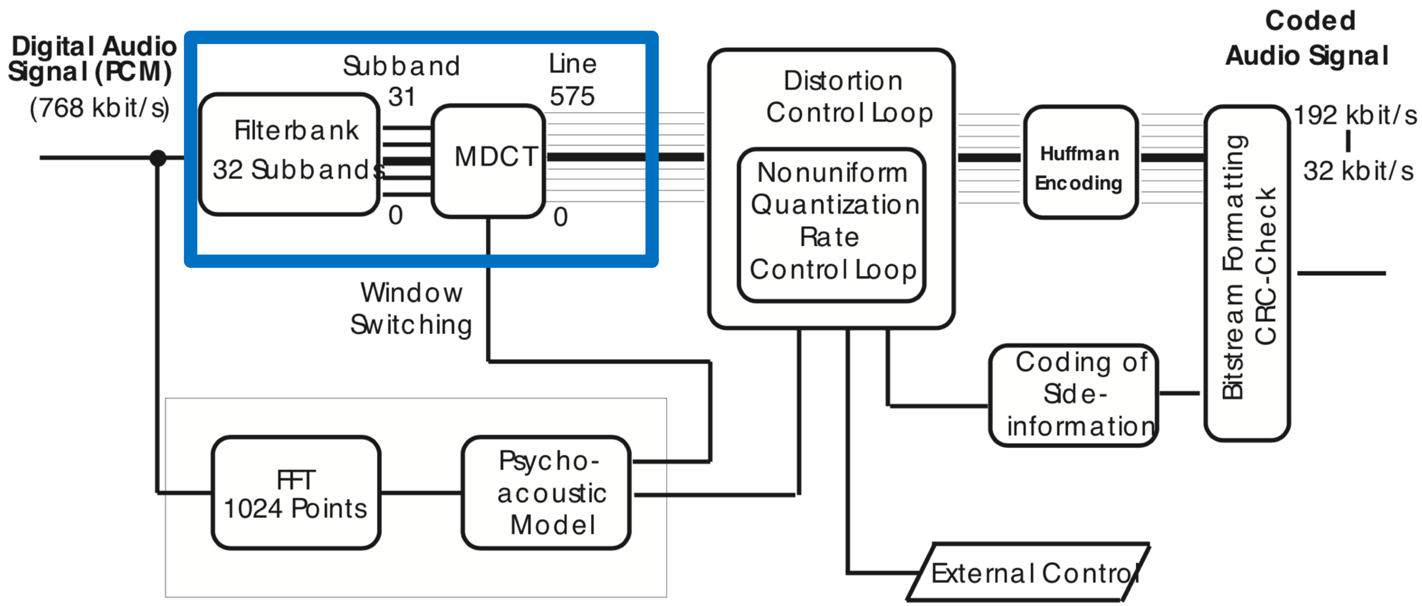
Output : 32 equally spaced frequency subbands

Ex) Suppose the sample freq. of the PCM signal is 44.1KHz.

Then Nyquist freq. is 22.05KHz, and each subband will be approximately 22050/32 = 689Hz wide.

So the lowest subband have a range from 0-688Hz, the next subband 689-1378Hz, etc.

MPEG-1 Audio Layer-3 Encoder



Modified Discrete Cosine Transform (MDCT)

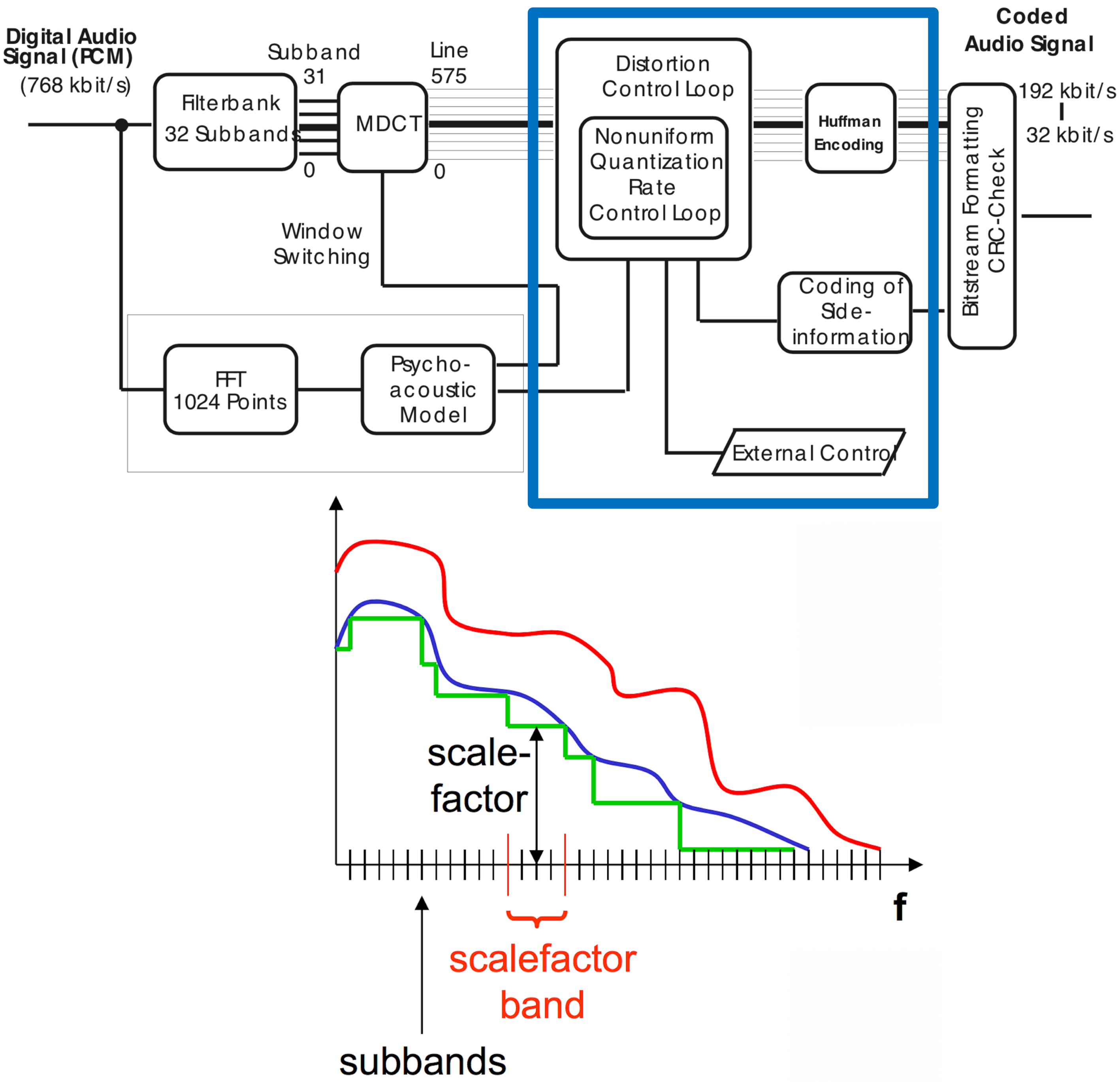
Input : 32 equally spaced frequency subbands

A sequence of window functions

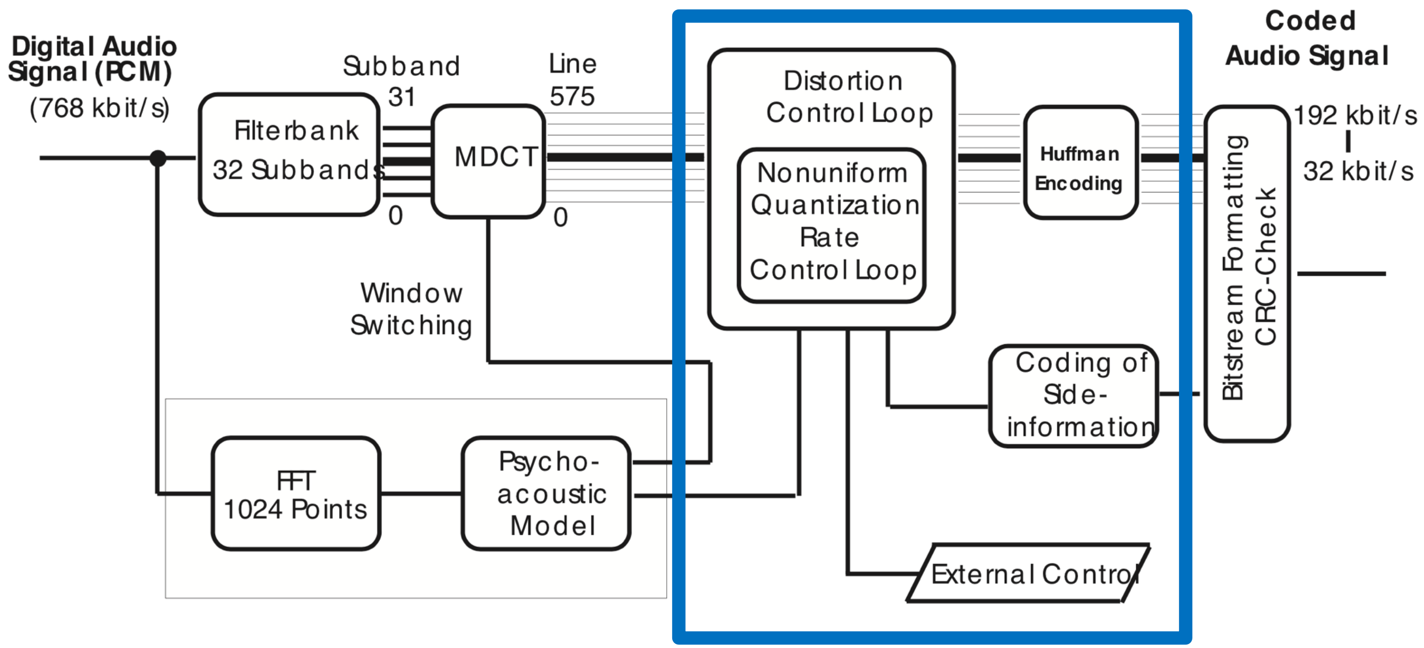
Output : 576 frequency lines (Each subband is split into 18 finer subbands)

- Increases the potential for redundancy removal.
- The error signal can be controlled to allow a finer tracking of the masking threshold.

MPEG-1 Audio Layer-3 Encoder



MPEG-1 Audio Layer-3 Encoder



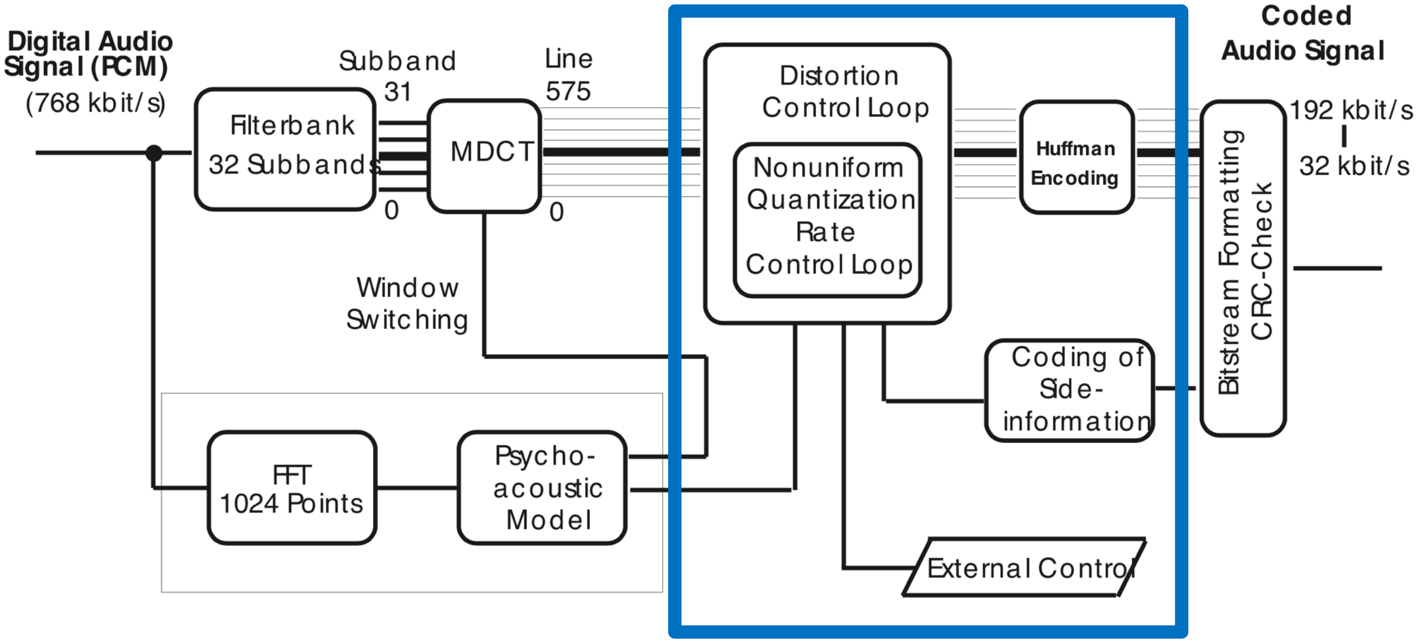
Input :

- Vector of the magnitudes of the 576 spectral values
- The number of scalefactor bands
- The allowed noise of the scalefactor bands
- Bits available for the Huffman coding and the coding of the scalefactors

Output :

- Vector of 576 quantized values (encoded)
- The scalefactors and quantizer step size information
- Huffman code related side information

MPEG-1 Audio Layer-3 Encoder



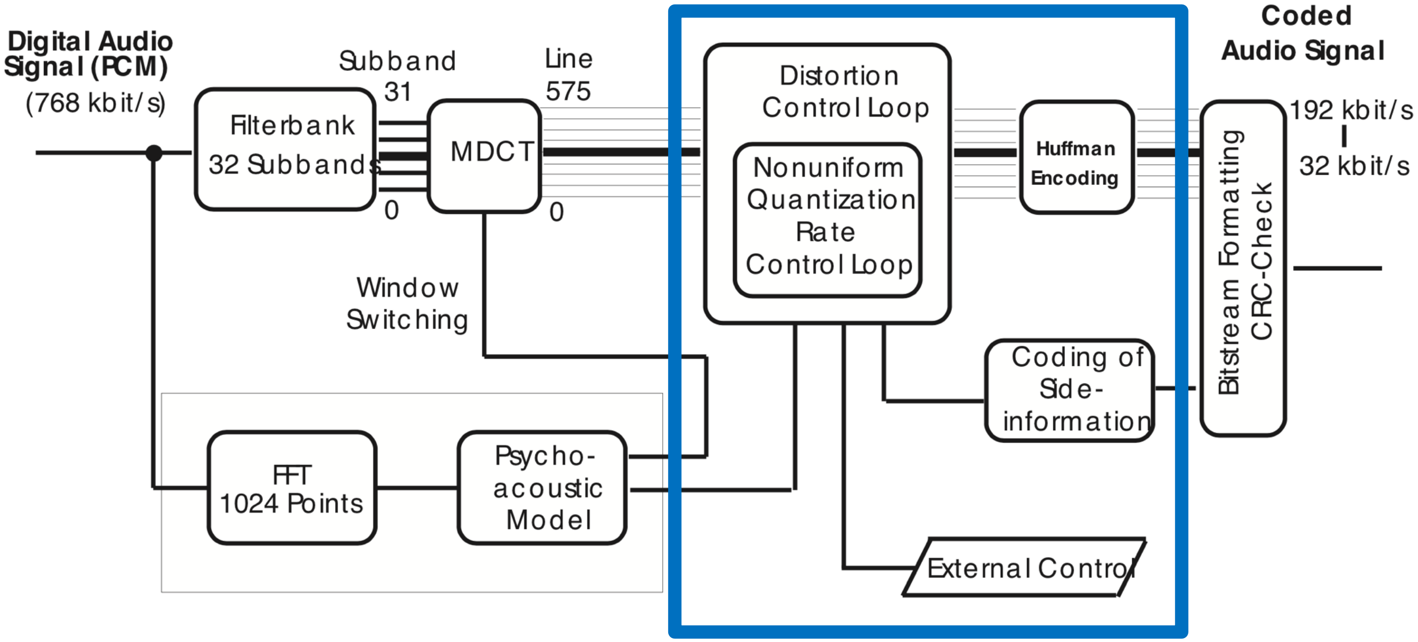
Inner Loop (Rate Loop)

- The samples are quantized with an increasing step size until the quantized values can be coded using one of the available Huffman code tables

- (Fixed) Huffman code tables assign shorter code words to more frequent quantized values.

- If (# of bits resulting from coding) > (# of bits available), then the entire procedure is repeated until the available bits are sufficient.

MPEG-1 Audio Layer-3 Encoder



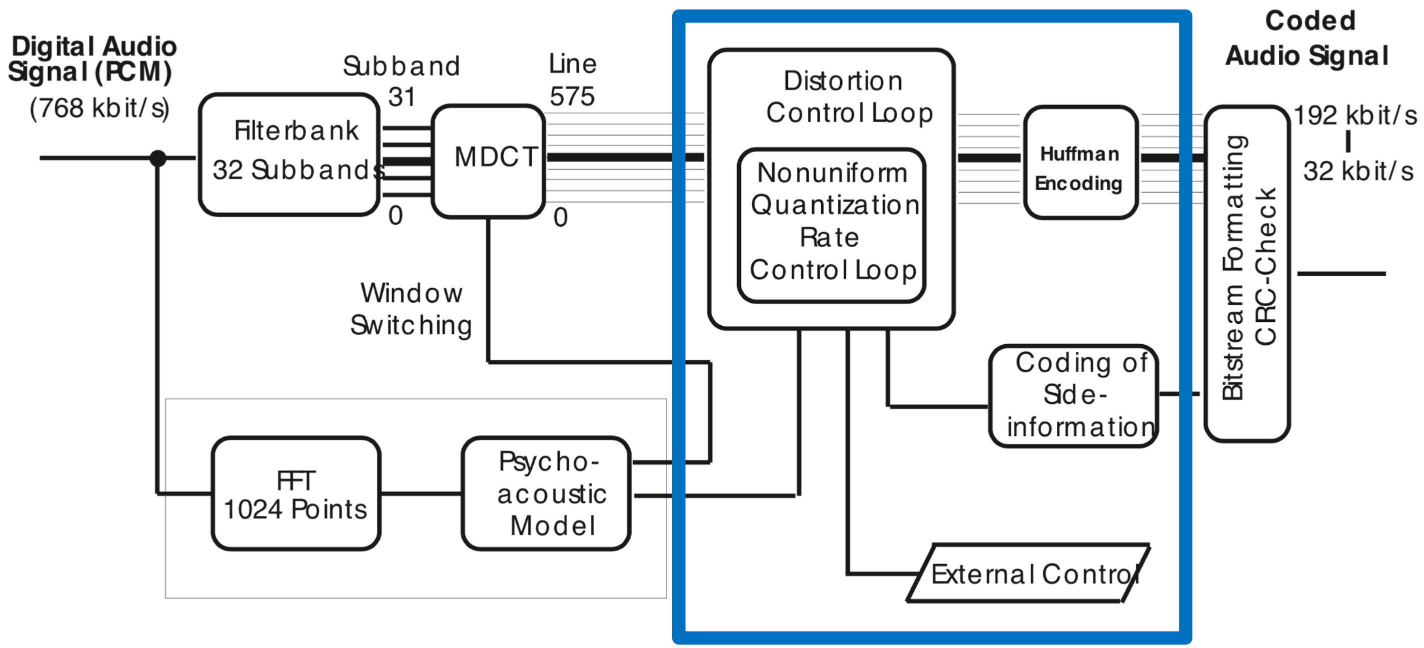
Outer Loop (Noise Control Loop)

- Scalefactors are applied to the frequency lines within each scalefactor band to shape the quantization noise.

- Inner loop is called and it returns 576 encoded quantized values

- If there exist scalefactor bands with noise exceeding the threshold, then increase the value scalefactors belonging to bands that are too noisy and repeat the entire procedure.

MPEG-1 Audio Layer-3 Encoder



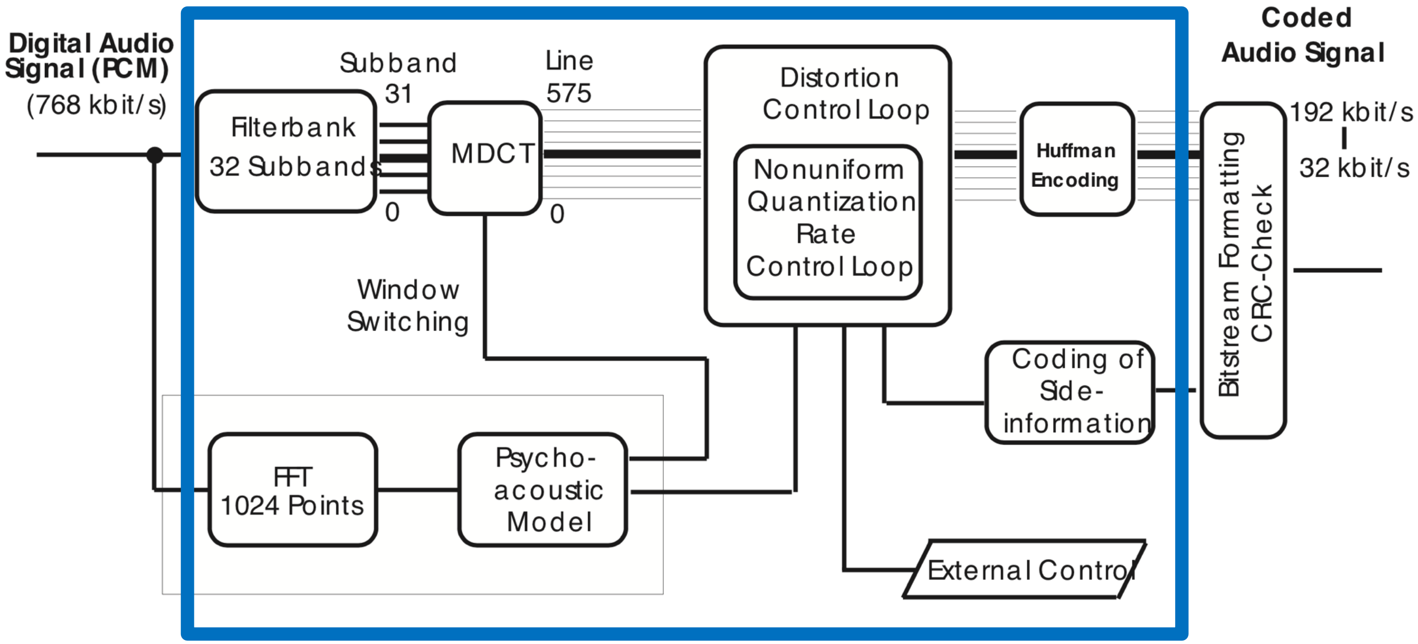
Ex) Suppose uncompressed scalefactor band information is represented by the number 12592.

- Quantizing it by dividing by 100 with a scalefactor of 1.0 returns 126. Restoring it returns 12600 and it differs from the original by 8.

- Quantizing it by dividing by 1000 with a scalefactor of 1.0 returns 13. Restoring it returns 13000 and it differs from the original by 408.

- Quantizing it by dividing by 1000 with a scalefactor of 2.0 returns 25. Restoring it returns 12500 and it differs from the original by 92.

MPEG-1 Audio Layer-3 Encoder



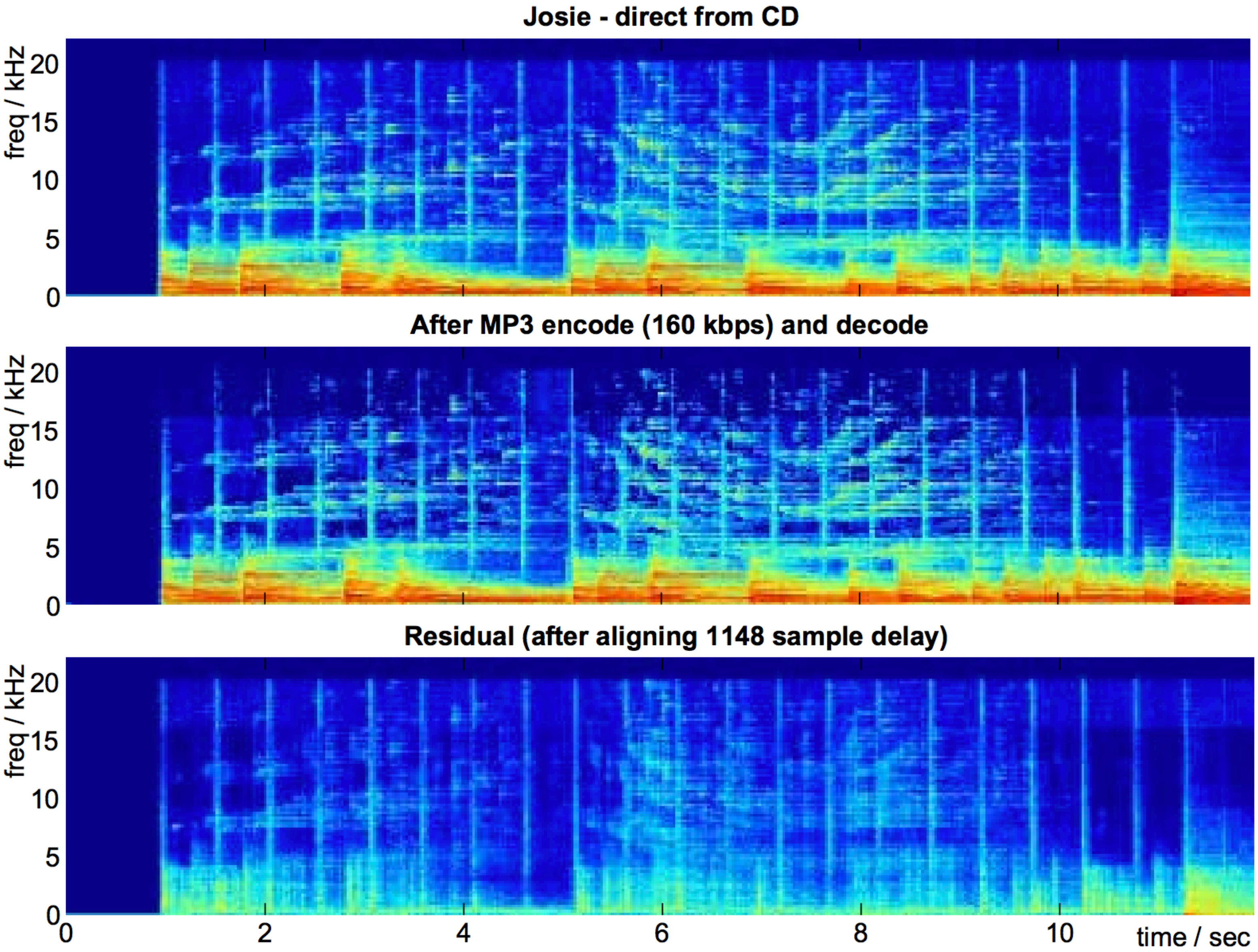
Input :

- A sequence of 1152 PCM samples

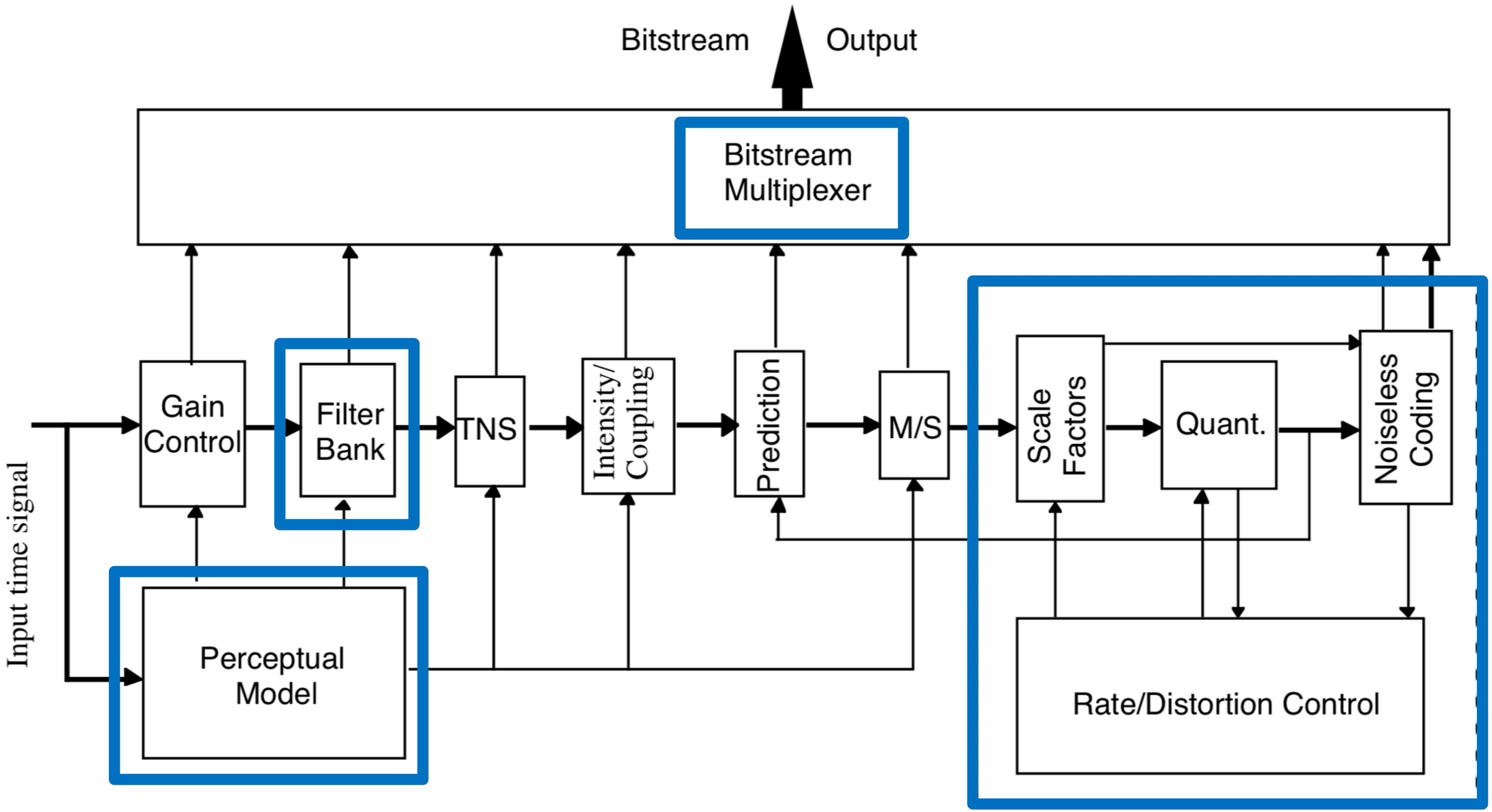
Output :

- Vector of 576 quantized values (encoded)
- The scalefactors and quantizer step size information
- Huffman code related side information

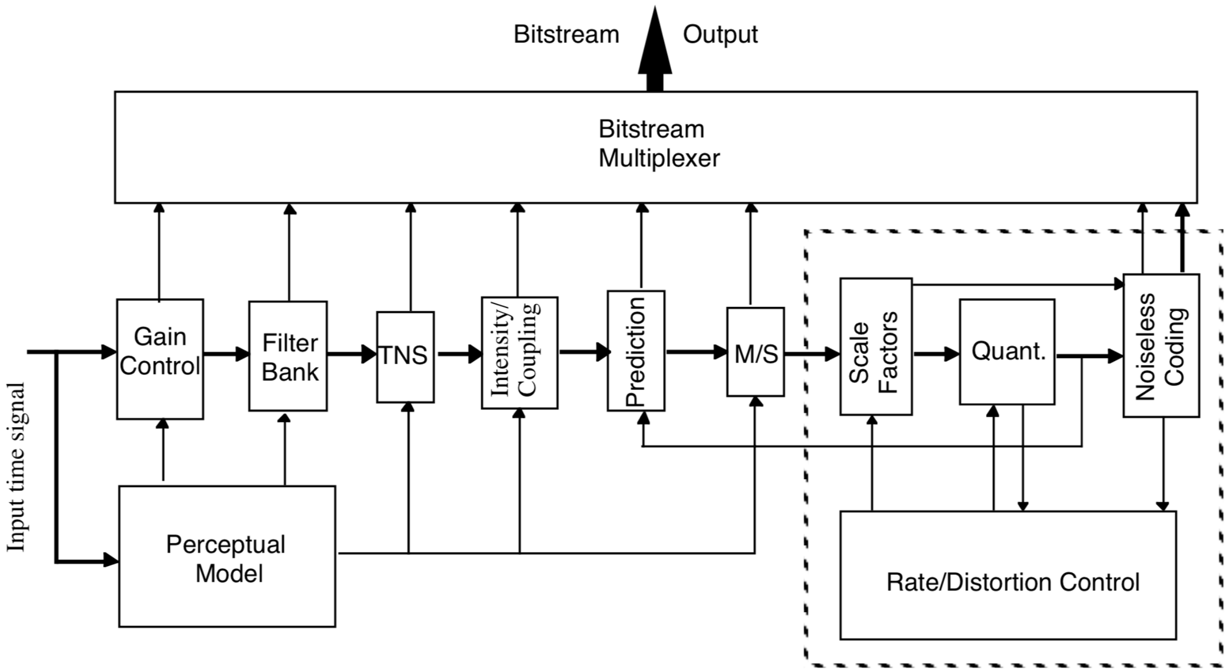
The Effects of MP3 Coding



MPEG-2 Advanced Audio Coding



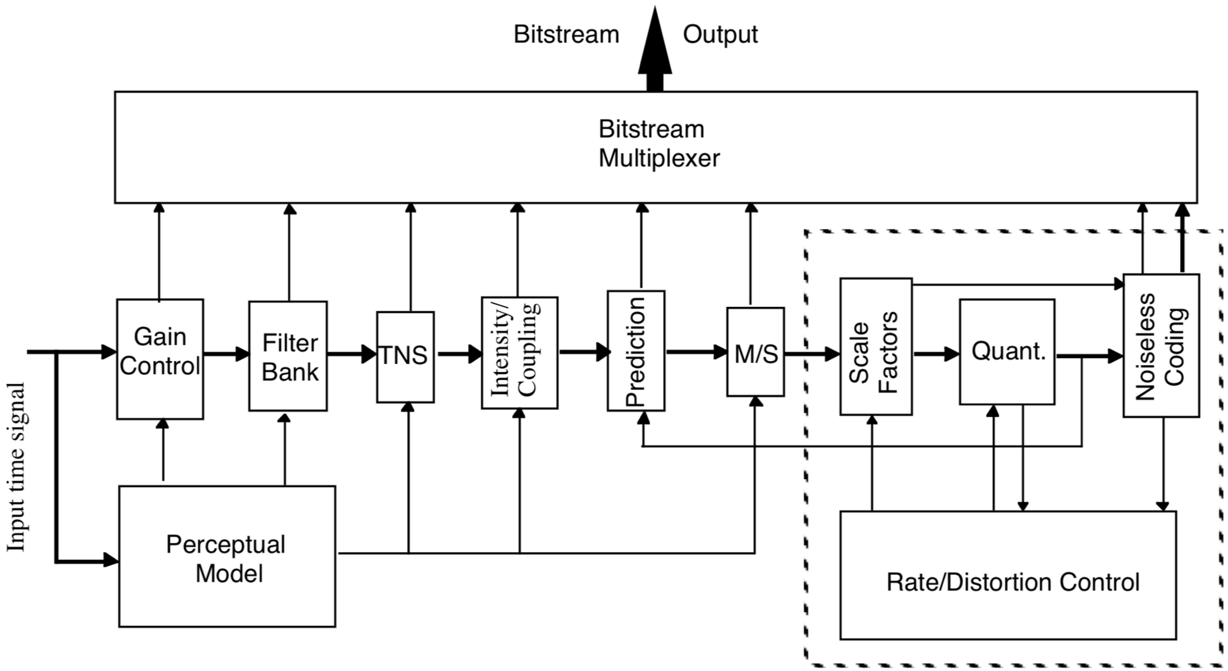
MPEG-2 Advanced Audio Coding



Tools to Enhance Coding Efficiency

- # of frequency lines in AAC is up to 1024 compared to 576 for MP3
- A prediction module improves the performance of the quantizer in cases where the original audio features patterns, such as high tonality
- Joint stereo coding is improved, allowing to reduce the bit-rate more frequently
- Huffman coding by quadruples of frequency lines is applied more often

MPEG-2 Advanced Audio Coding



Tools to Enhance Audio Quality

- Temporal noise shaping (TNS) minimizes the effect of temporal spread. This improves mostly the quantization (and hence the compression) of voice signals.

Compression Ratio of MP3 and AAC

Coding	Ratio	Required Bitrate
PCM CD Quality	1:1	1.4Mbps
MPEG-1 Audio Layer-1	4:1	384Kbps
MPEG-1 Audio Layer-2	8:1	192Kbps
MPEG-1 Audio Layer-3	12:1	128Kbps
MPEG-2 Advanced Audio Coding	16:1	96Kbps

Table 1. Bitrate required to transmit a CD quality stereo signal

Thank You

References

http://www.grc.upv.es/docencia/tra/referencias/AudioCoding/Brandenburg_mp3_aac.pdf

https://www.mp3-tech.org/programmer/docs/mp3_theory.pdf

https://www.tu-ilmenau.de/fileadmin/media/mt/lehre/ma_mt/audiocoding/Lecture_WS2013_14/06_bdg_Quantization_WS2013-14.pdf

https://www.ee.columbia.edu/~dpwe/e6820/lectures/E6820-L07-coding-2up.pdf

https://arstechnica.com/features/2007/10/the-audiofile-understanding-mp3-compression/

https://ee.stanford.edu/~osgood/Sophomore%20College/Audio%20Compression%20and%20the%20MP3%20Standard.pdf